

Natural language processing

From text to data.

by Pokey Rule

What is NLP?

“ **Natural language processing (NLP)** is a field of computer science, artificial intelligence and computational linguistics concerned with the interactions between computers and human (natural) languages, and, in particular, concerned with programming computers to fruitfully process large natural language corpora.

– *Wikipedia*

Areas of NLP

Syntax	Semantics	Discourse	Speech
Lemmatization	Machine translation	Automatic summarization	Speech recognition
Part-of-speech tagging	Named entity recognition (NER)	Coreference resolution	Speech segmentation
Parsing	Natural language generation	Discourse analysis	Text-to-speech
Sentence breaking	Natural language understanding		
Word segmentation	Question answering		
	Relation extraction		
	Sentiment analysis		
	Topic recognition		
	Word sense disambiguation		

Source: [Wikipedia](#)

NLP at Globality



Provider network growth

Automatically discover new service providers and infer their location, service type, and industry experience



Dynamic Q&A (Onboarding/Project Brief)

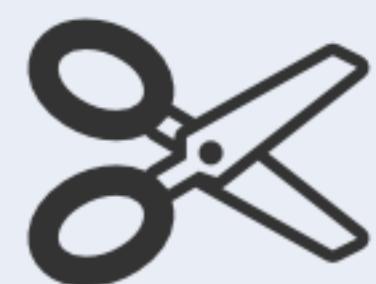
Infer information about a client or service provider based on free-form textual input



Conversational agents

Chat bots can automate common user interactions.

Pipeline



01. Tokenization

Split the text into a sequence of separate words

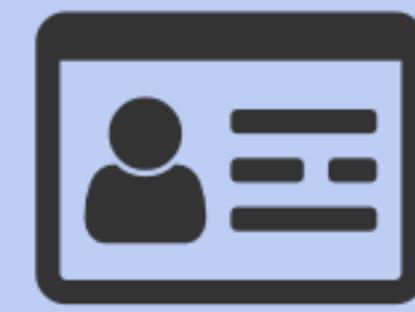
```
[ "We", "are", "located",  
"in", "Germany" ]
```



02. POS tagging

Determine the part of speech of each word

We	are	located	in	Germany.
PRP	VBP	VBN	IN	NNP



03. Named entity recognition

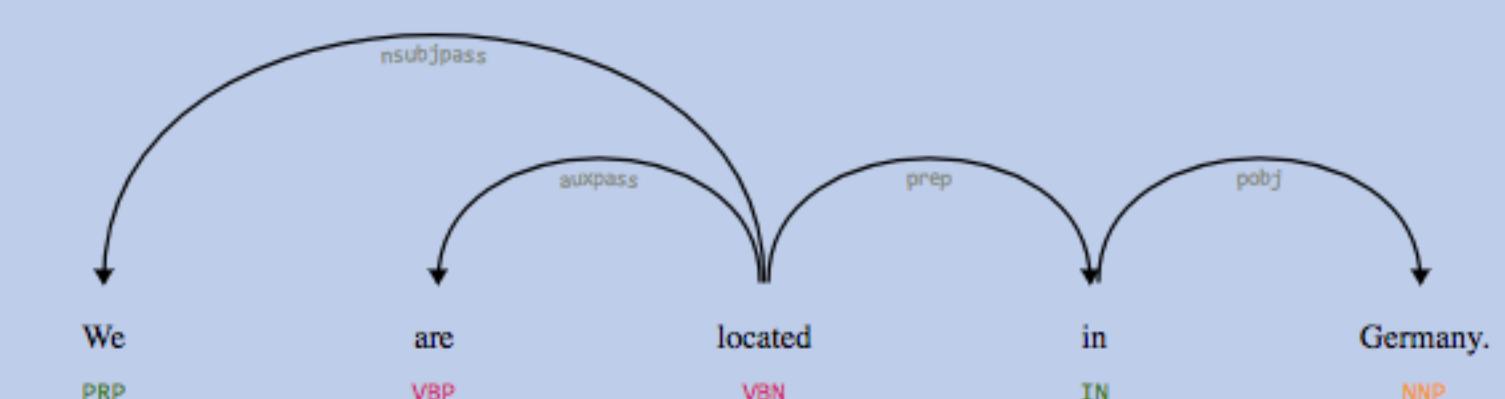
Identify real-world entities mentioned in text

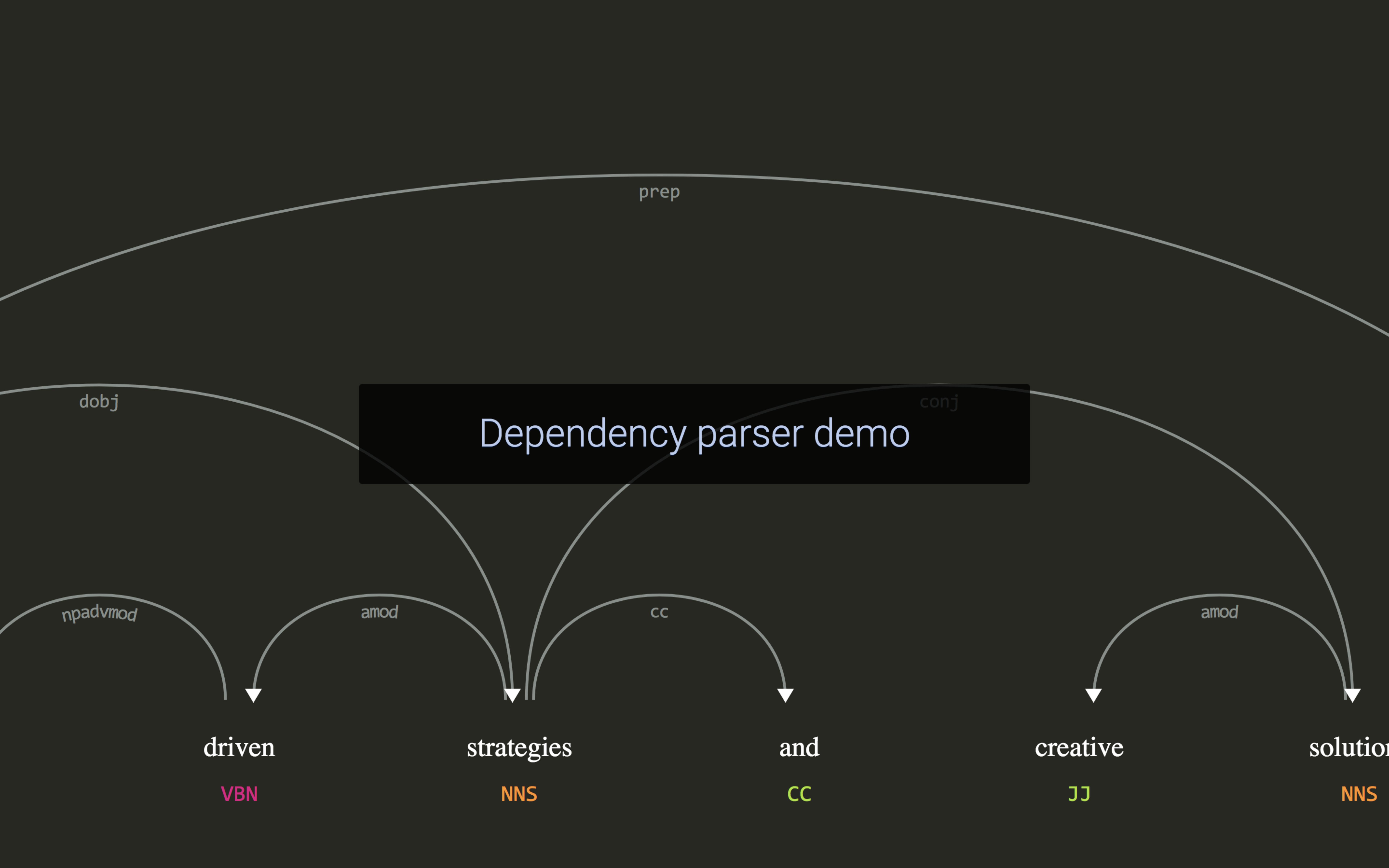
Germany

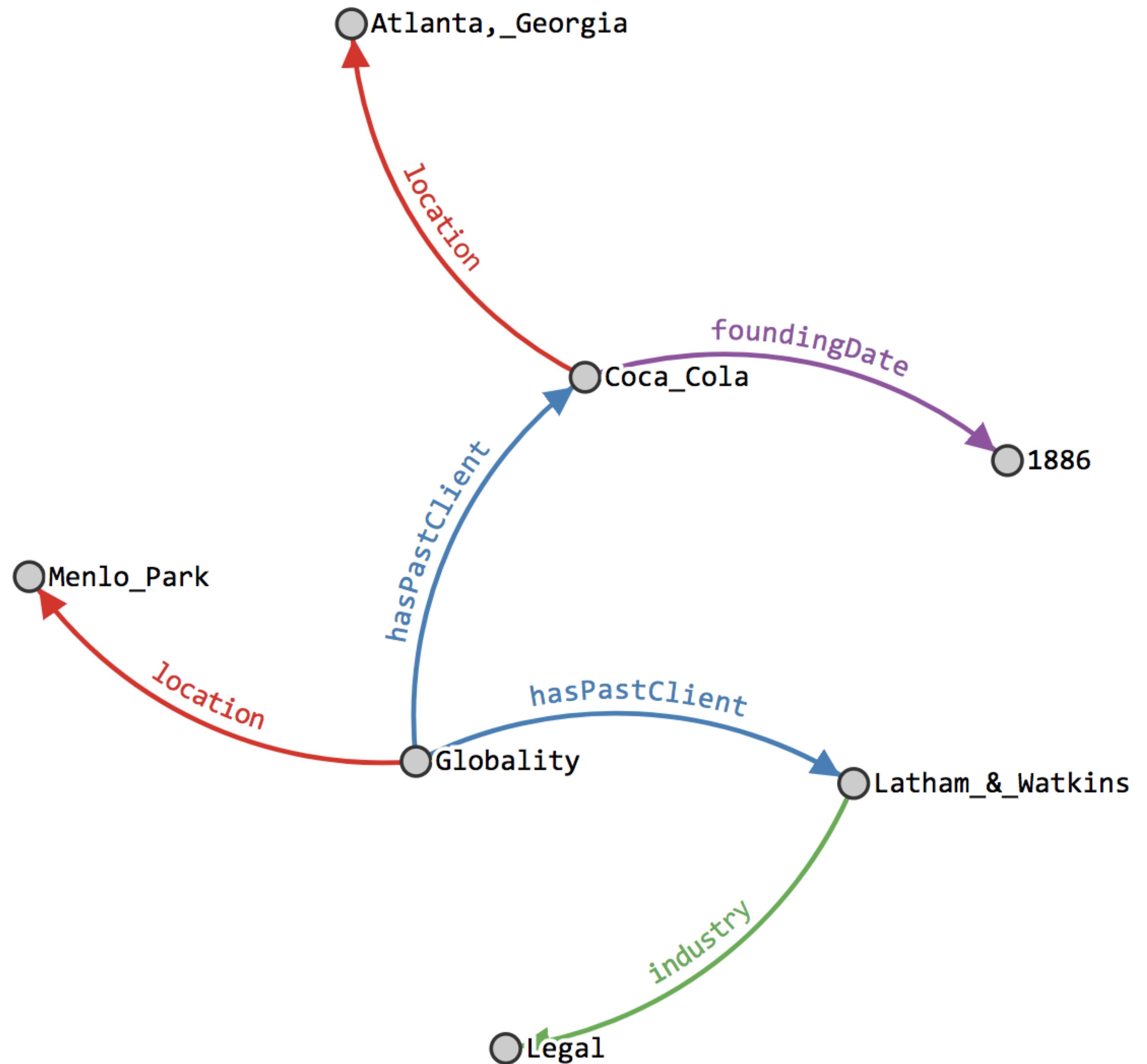


04. Dependency parsing

Parse the sentence into a hierarchical structure





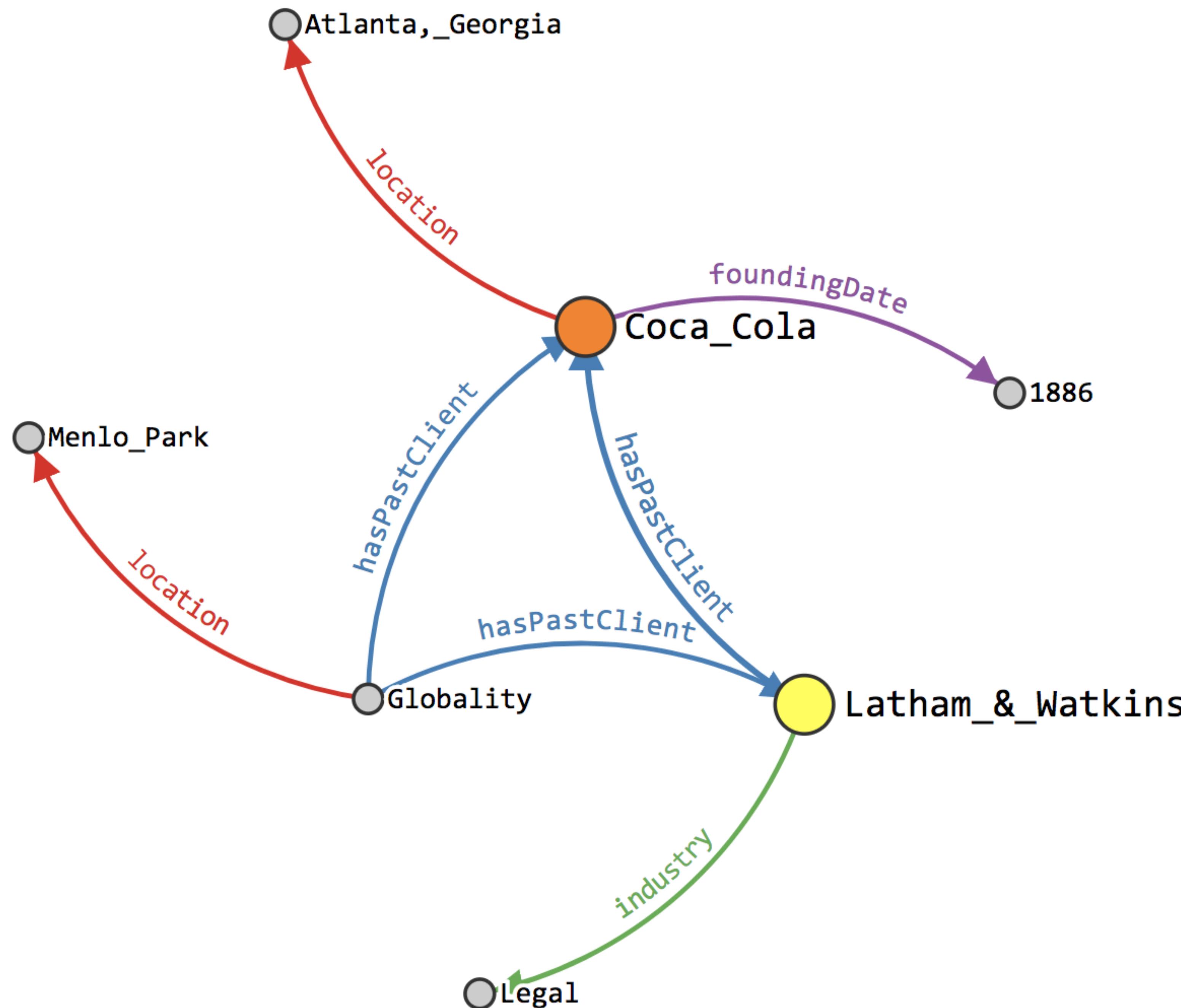


Knowledge graph

Knowledge graph is a collection of relations between entities.

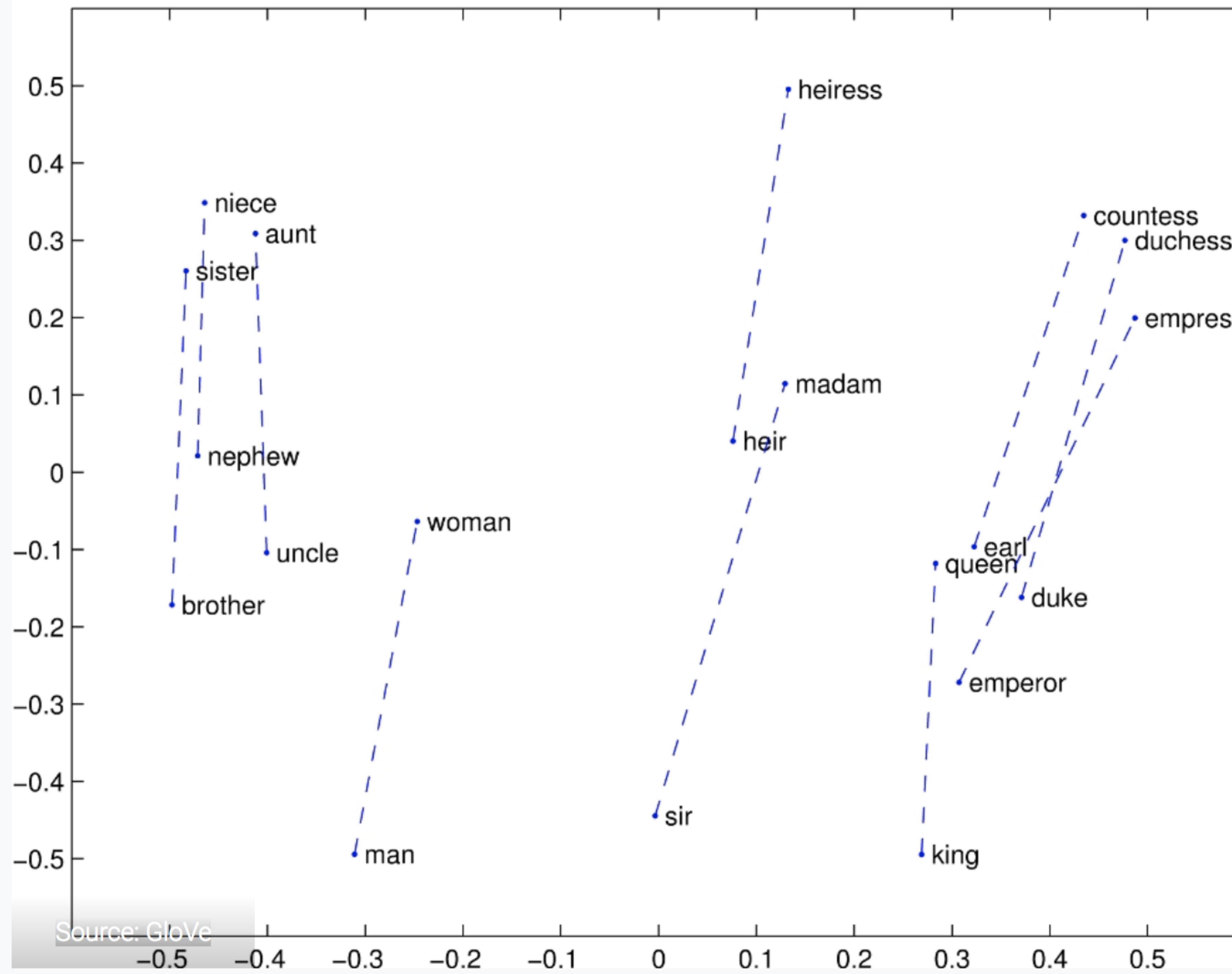
Contains information useful for matching

Given a sentence, we'd like to find relevant entities in knowledge graph and add new relations



Relation extraction

Latham & Watkins has had many high-profile clients, including
Coca-Cola.



Word vectors

Represent words as high-dimensional vectors

Learn dimensions from co-occurrences in real text

Dimensions capture semantic relationships

squad
regular-season

football
varsity

task

medal

improved

unit

undefeated

champions

semi-final

qualifiers

vault

qualifying

group

helped
hopes

bid

talented

Word embedding demo

teamed

joined
friend

mentor

youths

school

headquarters

accused

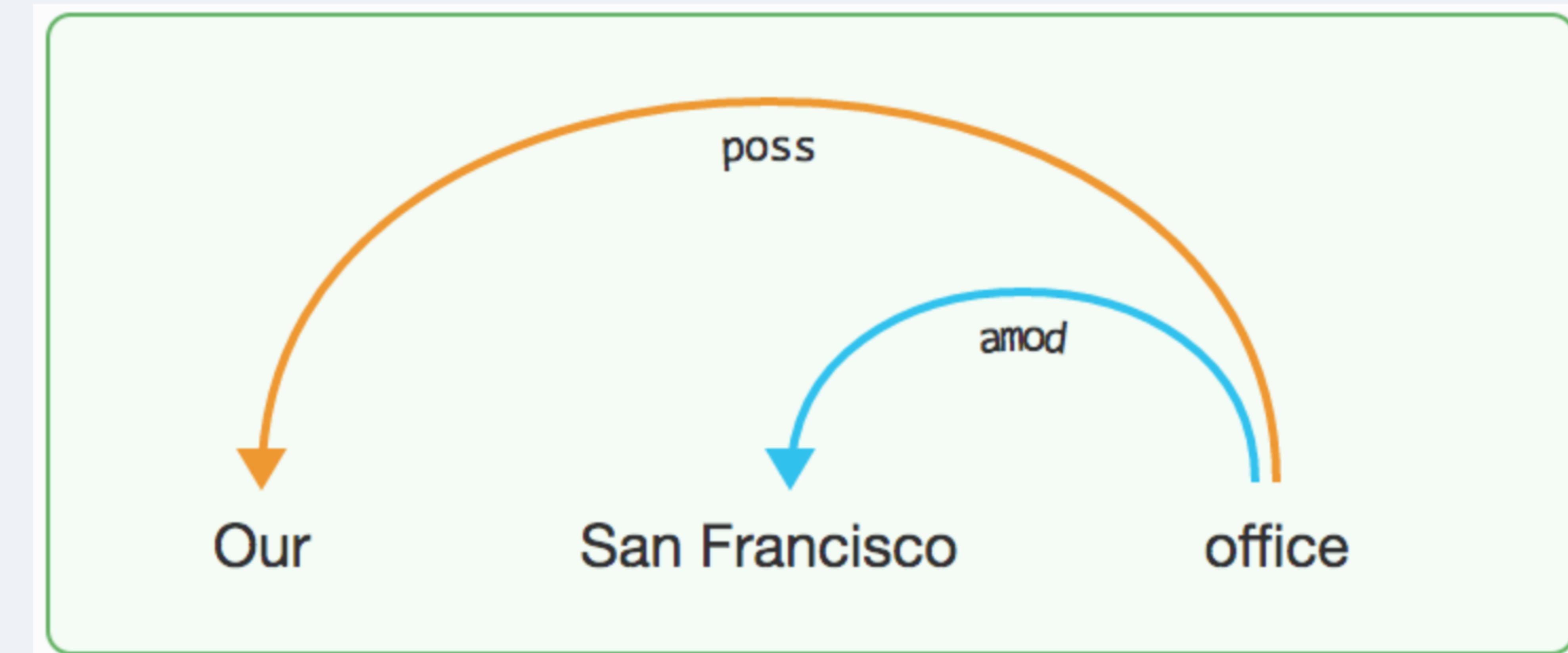
officers

recruits

camps

official
assistant

Relation extraction demo



Thank you.